

Dave Hamel is the Product Manager for Web Analytics, SEO, and the Google Search Engine here at CBC/Radio-Canada, and he has been working on the Web for over fifteen years. He has worked on sites for clients such as Citibank, Rolex, and Thankyou Network. Dave is an avid cyclist and a member of the Lapdogs cycling club.

There was a time when, if you wanted to look something up, you needed to go to a library or resort to a dusty collection of encyclopaedias to check a fact. The Internet changed that by literally placing information at our fingertips. The Internet had a problem though: with such vast amounts of information available, how could it make that information searchable? In the early days, there were directories such as Yahoo, Excite, and DMOZ, similar to the yellow pages in the phonebook, where websites would be analysed and categorised, but marketers soon learned to “game” these directories for more traffic, which caused the value and relevancy of these tools to decline.

Then along came Google and everything changed. Having performed over 65% of English language searches (*Source: comScore qSearch, February 2011*), rather than relying on websites to tell the search engine where they belonged, Google changed the game by telling the websites where they belonged instead. This process became known as “organic”, “natural”, or “algorithmic” search. This changed the way in which information was collected and left many people to ask such questions as:

How does Google decide which site is number one?

Why is my website not showing up?

How do I improve my position in the search results?

Luckily, Google itself provides the answer, and it is a process called **Search Engine Optimization**. SEO is the practice of changing a website to improve its position on the search results page. This is not strictly an IT practice either; it could involve coding, but it might consist of changing the copy or title, improving the user experience by changing the hierarchy, or improving the website’s marketing and PR. If you touch a website at all, SEO is your concern and it boils down to one simple thing: **The best SEO is to make your content worth sharing.**

How do you know what the best restaurant in town is? Why did you take your car to be fixed at that garage? Why did you go see that Indie film instead of the Hollywood one? Chances are that your decision was made partly because someone told you about it. SEO can be thought of as “word-of-mouth” for the Internet and it matters because there are over 270 million unique websites, over two billion people on the Internet, and Google alone performs over a billion searches a day! As such, you have to ask yourself how anyone is going to find your site.

SEO also matters because top search results receive 40% of the clicks from users, second search results receive 10%, third search results receive only 8.5%, and the percentage of clicks declines quickly from that point onward, which means that, if you are not on the first page of results, chances are that no one is finding you. However, search has a massive long tail. Over 70% of searches are for words searched five times or less per month. This means that you do not need to be number one for generic searches to be successful. For example, you might not be able to get to the top spot for “beans”, but you might manage to get the top spot for “Peruvian lima beans”.

The question then becomes, “How does it work?”

Google sends out a spider, or a “bot”, which crawls the Internet and collects information. This information is then indexed and an algorithm is applied to determine the importance of the page. The algorithm takes a number of things into account when assigning a pagerank (named after Google founder Larry Page). Pagerank is the numeric representation of the perceived value of a Web page in relation to the Internet, and it is made up of a number of different elements. These include:

- The authority of the host domain,
- The link popularity of a page,
- The anchor text of the external links,
- On-page keyword usage,
- Registration and hosting data,
- Traffic and click-through rate, and
- Social media metrics.

Domain authority is easily explained as how trustworthy or relevant a site is when compared to other sites. Essentially, what is the perceived importance for end users? For example, who are you going to trust more, www.apple.com or www.win-ipad-gamble.tv? The Internet has already decided and applied more importance to the Apple domain. However, domain authority is complex and there are over 150 different elements used to calculate this value, so it is very difficult to influence directly.

Link popularity is simply how many links are there to a specific page. Think about it this way, each link is a vote for a webpage, the more links, the more votes. This is combined with the anchor text of these links. “Ontario government doesn’t know if smart meters are working” is pretty descriptive link text. You can tell what the content will be

if you click on this link. Something like “[click here](#)” is not descriptive at all and could be a link to anything. As a result, Google will apply more weight to the first link.

If you have a variety of links that all have similar text, then Google assigns more credibility to that landing page. Therefore, using the previous example, if there is another link that reads “[Smart meters broken, claims Ontario government](#)”, the Google algorithm will recognise that page as being an authority on the subject of smart meters and the Ontario government. It does this by following the link to its destination and reading the content of the page.

Specifically, it looks at **the title tag**, which is used as the text in bookmarks, the tab information in your browser, and the link text for Google itself. It also looks for an *h1* or *header* tag, like a story headline and the actual body of content itself.

This leads us to the next aspect of page rank, which is **keyword density** or **on-page keyword usage**. The spider has already picked up the keywords from the link; in our example, these include “smart meter”, “Ontario”, and “government”. It will then read the copy on the page to ensure that these words are used again and that the page is indeed about what the link copy said it should be about.

If the link went to a page about monkeys instead of smart meters, Google would discount the link. If there were enough of them, it would start to penalise the page, since it would assume someone was trying to scam the system. The algorithm also takes readability into account. In other words, you cannot pepper the page endlessly with the words “smart meter”. You need to write the copy with the end user in mind.

Registration and hosting data work as part of SEO as well. Google places more weight on a “.com” domain than on a country-specific domain such as “.ca”. Furthermore, since “.com” has been around longer, there is a good chance that, if you have a one-word domain with “.com”, you have had it for a long period of time, which also increases the pagerank. The algorithm takes into account how long your site has been at its current domain, how many other sites belong to that domain, and what other domains are associated with your company.

URLs play a small role in SEO as well. A URL like www.cbc.ca/news/ontario-government-smart-meters-useless is better than www.cbc.ca/n/?postid=536. Since the URL itself contains information on what the content is likely to be, Google will reward this effort.

Traffic and click-through rate are Google’s criteria when it comes to checking its own results to influence what users are seeking. For example, if you searched for “smart meters” and two links came up with comparable pagerank but more people clicked on the

second link, Google would eventually promote that link to number one, recognising that it had a higher response rate and was probably the more relevant result.

Lastly, with the advent of new **social media** technologies such as Facebook and Twitter, not all links are on websites *per se* or surrounded by large bodies of copy. However, these links are important since social media can reach a large number of people. As a result, the Google algorithm includes a social media aspect, although its importance is not that great just yet.

The weight or importance of each of these values changes as Google constantly tweaks its algorithm, so no one really knows with certainty how many Twitter links are equal to a website link, but we do know that those from respected sites or sites with a high pagerank count more than those from sites with low pagerank. In simple terms, a link from CBC/Radio-Canada is worth more than a link from Davesblog.blogspot.com. Some people attempt to use, “*grey techniques*” such as link farms or gateway pages to influence Google in their favour. Techniques that may have worked in the past no longer work today and may be penalised tomorrow. As such, trying to trick Google will only hurt your efforts in the long run.

Now that we know what Google considers important, what can you do to improve your search ranking?

Rule #1 - Start with the basics: Avoid worrying about viral marketing or link-building partnerships until you have ensured that you have accessible, unique, quality content with specific, targeted keywords. Otherwise, you are just attempting to build upon a poor foundation. Remember that searching is actually about people and not about websites. Therefore, try to think of your end users first when working on your website.

Rule #2 - Quality content only: If you create content, make sure it is relevant and user facing. **Only** create content that solves problems, answers questions, and provides information. Consider this: Why should a third party link to your content? Why would they tell anyone else? Look at this webpage (<http://www.cbc.ca/stevenandchris/2011/07/seven-countertop-wine-racks.html>) and ask yourself what you could write about it to tell others and entice people to come look at it? What problems does it solve or information does it provide?

Rule #3 – Try to use four to six keywords specific to that page: Content performs better in search engines when it is targeted for specific searches. You can look at the current metrics to see what words people are using to find your content. For example, if your customers are searching for “running shoes” to find you, don’t label your content “sneakers”; use the words “running shoes” and reinforce those keywords.

Rule #4 – Search engines cannot “see” images, so ensure that your images have *alt* tags. This not only improves the SEO, it also makes your website more accessible. You should also ensure that your code is well formed and W3C validated (<http://validator.w3.org/>). These details require small amounts of effort and show Google that you actually care about your end users.

Rule #5 – Write a smart *title* tag: It is the text for the link in Google search results, the title bar on the Web browser, and the text in the bookmark, so its importance cannot be understated. For example, if your page is about Peruvian lima beans, you might have a title like “Peruvian lima beans, the nutritious Incan food from South America” the idea being you could catch traffic for the terms “Peruvian”, “lima”, “beans”, “nutritious”, “Incan”, and “South America”.

Rule #6 – Write an informative meta-description: These are used by Google as the description for the link and are essentially your elevator pitch to get someone to visit your site. Tell users what they will find if they click on the link.

Rule #7 – Respect the SEO cycle: People will read your content and, if it is compelling, they will link to it. The more people link to it, the better the results in the search engine index will be. The better the result, the more people will find your quality content. **The cycle repeats itself and SEO happiness ensues.**

The screenshot shows a web browser displaying a CBC News article. Red boxes highlight the following text:

- Browser tab: **Giant tuna caught in Tokyo fetches record \$747K**
- Address bar: **www.cbc.ca/news/technology/story/2012/01/05/tuna-tokyo-record.html**
- Article headline: **Giant tuna caught in Tokyo fetches record \$747K**
- Sub-headline: **Winning bidder at auction heads sushi restaurant chain**
- Video player title: **\$747K tuna 0:53**
- Video player description: **A bluefin tuna caught off northeastern Japan fetched a record 56.49 million yen (about \$746,797 Cdn) on Thursday in the first auction of the year at Tokyo's Tsukiji fish market.**

CBC/Radio-Canada does many of these things correctly, but there is still room for improvement. The story depicted above has “Giant tuna caught in Tokyo fetches record \$747K” as its headline. The keywords are “giant”, “tuna”, “caught”, “Tokyo”, “fetches”, and “record”; therefore, these are the five keywords that we want to see throughout our page. If we start with the URL, we can see “tuna-tokyo-record.html”, and it contains only 50% of the keywords. There is room for improvement in that.

The *title* tag, represented by the tab text in the top left, the bookmark text, and the highlight text contain all the keywords followed by hierarchical information, the section, and the website. That is good.

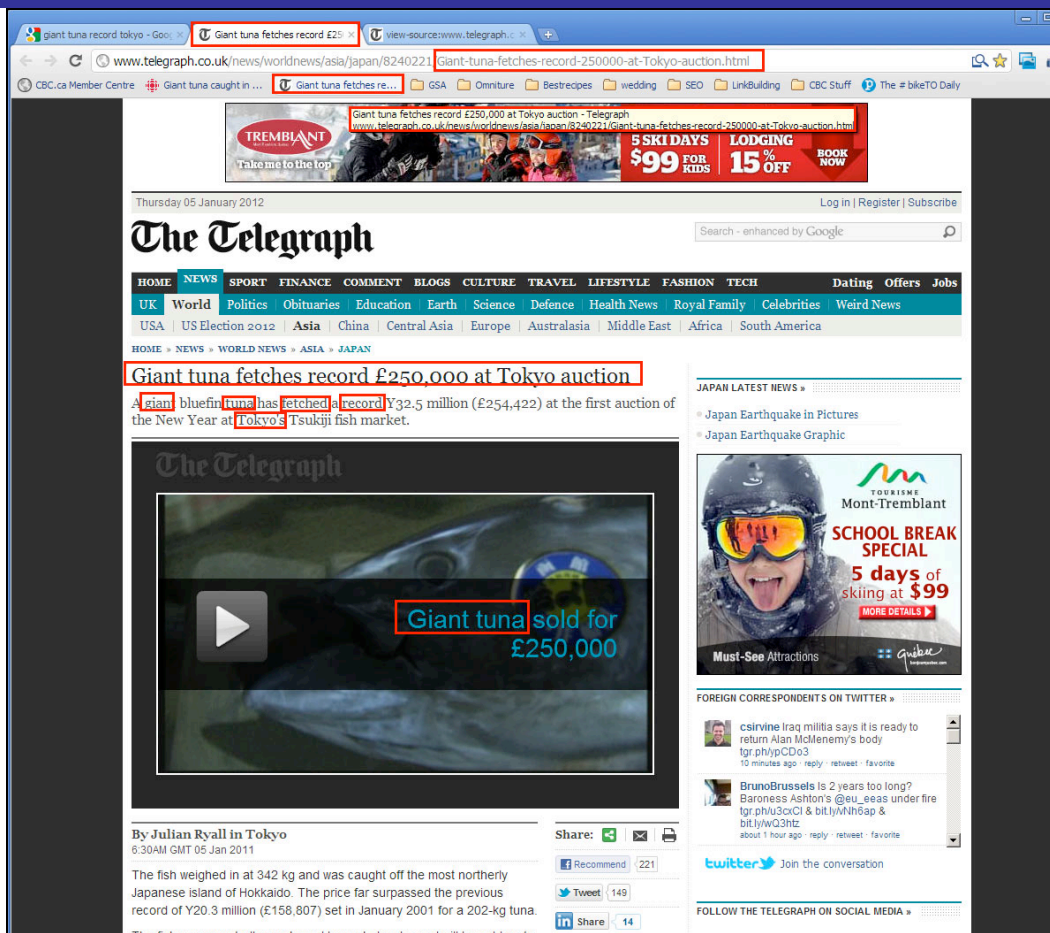
The copy at the start of the article contains “tuna”, “caught”, “fetch”, “Tokyo”, and “record”, but it is missing the word “giant”. By using all the keywords, it helps reinforce the concept of what this page is about, namely a giant tuna fish that was caught and fetched a record price in Tokyo. *See what I did there?*

When we look at the result in Google, we find more information on how we can improve our search results and further optimise our website.

[Tuna caught in Tokyo fetches record \\$747K - CBC News](http://www.cbc.ca/news/offbeat/story/2012/01/05/tuna-tokyo-record.html)
www.cbc.ca/news/offbeat/story/2012/01/05/tuna-tokyo-record.html
29 minutes ago – A bluefin **tuna caught** off northeastern Japan **fetches** a record 56.49 million ... Skip to main content; Skip to **Tuna caught in Tokyo fetches** record \$747K ... U.S. election 2012 Iowa the **big** loser; what's in store for Romney CBC ...

The first thing to note is the difference in the URL that Google has compared to the one we looked at earlier. The latter fell under “technology”, whereas Google has it under “offbeat”, and this represents an issue. Anytime Google finds duplicate content, it picks a version to index and keep, and discards the other result. This means that, if bloggers link to the “technology” version of the story instead of the “offbeat” version and Google discarded the “technology” version, we would lose those link votes.

We can also see a bunch of text in the description that does not have anything to do with the story, namely the part at the end concerning the U.S. election. Upon looking at the code itself, it appears that this is intended for accessibility purposes; however, it is showing up where we do not want it.



When we search for “giant tuna fetches record Tokyo”, we see a page from The Telegraph in the UK come in first. This is partly because their page name in the URL includes all of those keywords: Giant-tuna-fetches-record-250000-at-Tokyo-auction.html. It also ranks first because they have a number of keywords repeated within the first sentence of the story.

The Telegraph page has also has its meta-description come right after the opening *header* tag, following the *title* tag, which may help in the SEO as well.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en"
xmlns:fb="http://www.facebook.com/2008/fbml" xmlns:og="http://opengraphprotocol.org/schema/">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />

<title>Giant tuna fetches record £250,000 at Tokyo auction - Telegraph</title>

<meta name="description" content="A giant bluefin tuna has fetched a record Y32.5 million
(&amp;pound;254,422) at the first auction of the New Year at Tokyo&#039;s Tsukiji fish market." />
<meta name="keywords" content="tuna, giant, tokyo, japan, auction, sushi, Japan,Asia,World
News,News" />
```

One key factor that we cannot see is the number of back links to the domain itself, the number of “votes” the site is getting. The Telegraph has roughly 1,110,000 back links, while CBC/Radio-Canada has a mere 638,000. The Telegraph page had 149 tweets about it, while the CBC/Radio-Canada one has three. Remember, word of mouth is key and links such as tweets are the Web’s version of word of mouth.

There are some technical issues; for example, CBC/Radio-Canada is also missing an XML sitemap to help guide the search engine to pages. It may not count for much, but every improvement helps.

In light of all this, some minor changes and a little effort may help improve our standing in the results. It also bears noting that some of the changes need to occur in the code, some in the hierarchy, and some in the copy itself.

Search Engine Optimisation is not a black art, and it is not magic. It is the patient process of making incremental improvements through better copy and more descriptive titles, through cleaner code and site architecture, through targeted keywords and thoughtful planning to improve your results.

Google worked hard to make search about returning relevant results and focusing on end users. If you do the same, Google will reward your efforts and, with a little patience, you will soon be number one on the results page.